学校编码：**10384** 分类号______密级 ______

学　　号：**X2005170006** **UDC______**

厦门大学

硕　士　学　位　论　文

双向变异S-粗集的概率特征

Probability character of two direction variation S-rough sets

修　明

指导教师姓名：林 亚 南 教 授

专 业 名 称：基 础 数 学

论文提交日期：2008 年 9 月

论文答辩时间：2008 年 月

学位授予日期：2008 年 月

答辩委员会主席：__________

评　阅　人：__________

2008 年 月

厦门大学学位论文原创性声明

兹呈交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

2008 年 9 月 30 日

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1、保密（ ），在 年解密后适用本授权书。

2、不保密（ ）。

（请在以上相应括号内打“√”）

作者签名： 日期：2008 年 9 月 30 日

导师签名： 日期：2008 年 9 月 30 日

目 录

Contents

摘 要

文献[1]提出了S-粗集，给出了它的两类结构：单向S-粗集，双向S-粗集．[2～16]对S-粗集的特性和应用，给出进一步讨论，推广了波兰数学家Z．Pawlak提出的粗集[17]．S-粗集与Z．Pawlak粗集的共同特征是以元素的R-等价类来定义的．S-粗集与Z．Pawlak粗集的主要差异是S-粗集具有动态性，而且S-粗集的动态性是通过元素迁移体现出来的，由于元素迁移在很多情况下是不确定的，表现出一定的随机性，也就是说不同的元素迁移发生的可能性大小是不同的．文献[18]在文献[1]的基础上提出变异S-粗集和它的结构，对S-粗集的本质拓宽了认识，由于S-粗集的随机特征，自然使得变异S-粗集也具有概率特征．在实际问题中，一个系统受到相关因素的随机干扰是普遍存在的，从变异S-粗集的角度来看就是属性迁移具有随机性．因此，作为粗系统分析的理论基础，变异S-粗集自然不能忽视属性迁移的随机性，在考虑属性迁移的随机性时，双向变异S-粗集有什么样的结构和特征？这一问题，在有关S-粗集的文献中尚未进行讨论．依据这一背景，本文提出了属性迁移的概率、概率函数的概念，讨论了双向变异S-粗集的概率特征．

关键词：属性迁移的随机性；属性迁移的概率；概率函数；依概率生成；随机结构

Abstract

S-rough sets, referring to Ref.[1], possess two kinds of structures, namely one direction S-rough sets and two direction S-rough sets. The more characteristics and applications of S-rough set are presented in refs.[2~16], which developed Pawlak rough sets theory [17] proposed by Poland mathematician Z.Pawlak. The common characteristic between S-rough set and Pawlak rough set is that they are all defined by element R-equivalence classes. The main different between S-rough sets and Pawlak rough sets is that S-rough set has dynamic characteristic which is embodied by element transfer. Since element transfer is uncertain mostly, behaving with some randomicity, namely the different element transfers occur with different probabilities. Variation S-rough set and its structure were proposed in refs.[18] on the basis of refs.[1], which extend the realization to the essence of S-rough set. For the randomicity of S-rough set, the variation S-rough set also has probability characteristic naturally. In practice problems, it is common that the system is interfered by associated factors randomly, in the other words, attribute transfer has randomicity at the viewpoint of variation S-rough set. So variation S-rough set, as the theoretical basis of rough system analyse, should not ignore the randomicity of attribute transfer naturally. What are the structure and characteristics of two direction variation S-rough set when the randomicity of attribute transfer is considered? This problem has not been researched in present literatures associated with S-rough sets theory. On this background, this dissertation proposed the concepts of the probability of attribute transfer and probability fuction, and discussed the probabilitic characteristic of two direction variation S-rough set.

Key Words: the randomicity of attribute transfer; the probability of attribute transfer; probability function; generation depending on probability; random structure

第一章　引言

当今，社会已经进入了网络信息时代，计算机与网络信息技术的飞速发展使得各个领域的数据和信息量急剧增加，并且由于人类的参与使数据与信息系统中的不确定性更加显著、更加复杂.如何从大量的、杂乱无章的、强干扰的海量数据中挖掘潜在的、有利用价值的信息（有用知识），这给人类的智能信息处理能力提出了前所未有的挑战，由此产生了人工智能研究的一个崭新领域——数据挖掘-知识发现（data mining-knowledge discovery）.在数据挖掘-知识发现诸多方法中，粗糙集理论与方法对于处理复杂系统不失为一种较为有效的方法，因为它与概率方法、模糊集方法和证据理论方法等其他处理不确定性问题理论的最显著的区别是它无需提供问题所需处理的数据集合之外的任何先验信息.

粗糙集理论是波兰数学家Z. Pawlak教授于1982年提出的一种数据分析理论.由于最初关于粗糙集理论的研究主要集中在波兰，因此当时并没有引起国际计算机界和数学界的重视，研究地域仅局限于东欧一些国家.直到1990年前后，由于该理论在数据的决策与分析、模式识别、机器学习与知识发现等方面的成功应用，才逐渐引起世界各国学者的广泛关注.1991年Z. Pawlak教授的专著《粗糙集——关于数据推理的理论》（Rough Sets——Theoretical Aspects of Reasoning about Data）的问世，标志着粗糙集理论及其应用的研究进入了活跃时期.1992年在波兰召开了关于粗糙集理论的第一届国际学术会议，1995年ACM Communication将粗糙集列为新浮现的计算机科学的研究课题.目前，粗糙集理论及其应用已成为信息科学最为活跃的研究领域之一.数据挖掘-

知识发现逐渐成为人们熟知的概念和研究领域，人们正在利用Z. Pawlak粗糙集这个新的数学工具，对数据挖掘-知识发现做深入研究，并获得一些令人兴奋的结果.

1982年波兰数学家Z. Pawlak教授给出粗糙集（以下简称粗集）的一般性研究，Z. Pawlak粗集是以R-元素等价类$[x]$定义的.在Z. Pawlak粗集中蕴含着这样的事实：元素集合$X \subset U$给定，元素等价关系R给定，则$X \subset U$的下近似$R_{-}(X)$确定，上近似$R^{-}(X)$确定，集合$X \subset U$的粗集（$R_{-}(X)$，$R^{-}(X)$）确定.这个事实意味着1°不允许集合X之外的元素x进入集合X（不允许等价类$[x]$之外的元素x进入等价类$[x]$），2°不允许集合X之内的元素x离开集合X（不允许等价类$[x]$之内的元素x离开等价类$[x]$）.显然Z. Pawlak粗集是一个具有静态特性集合$X \subset U$的粗集，这一特性限制了Z. Pawlak粗集的广泛应用.

利用Z. Pawlak粗集研究动态数据挖掘，动态知识发现遇到了困难.2002年山东大学史开泉教授对Z. Pawlak粗集做出改进，给出动态R-元素等价类$[x]$的概念，在文献[1]中提出了S-粗集（singular rough sets）. S-粗集是依赖下面的背景提出的：一个例子，设$[x]_A$，$[x]_B$，$[x]_C$是赴A，B，C三地旅游的团体，显然，$[x]_A$，$[x]_B$，$[x]_C$是关于旅游地A，B，C的等价类，$[x]_A$上的元素关于A是不可辨的，$[x]_B$上的元素关于B是不可辨的，$[x]_C$上的元素关于C是不可辨的.因为一些原因，$[x]_A$上的元素x_p，x_q取消了赴A地旅游的计划，则$[x]_A$的边界向内收缩，$[x]_A$具有单向动态特性. 因为一些原因，$[x]_B$之外的元素x_α，x_β，x_γ参入到$[x]_B$中，赴B地旅游，则$[x]_B$的边界向外扩张，$[x]_B$具有单向动

态特性. 因为一些原因，$[x]_C$ 中的元素 x_λ，x_k 取消了赴 C 地的旅游计划，$[x]_C$ 之外的元素 x_η，x_φ，x_σ，x_ε 参入到 $[x]_C$ 中，赴 C 地旅游，则 $[x]_C$ 的边界既向内收缩又向外扩张，$[x]_C$ 具有双向动态特性，显然 $[x]_A$，$[x]_B$，$[x]_C$ 具有的等价关系 R 没有改变，而包含 $[x]_A$，$[x]_B$，$[x]_C$ 的集合 $X \subset U$ 具有了动态特性：单向动态特性或者双向动态特性.文献[1]对S-粗集进行了讨论，文献[2～16]对S-粗集的特性和应用，给出了进一步讨论，推广了波兰数学家Z. Pawlak教授提出的粗集，S-粗集是以具有动态特性的 R-元素等价类[x]定义的， S-粗集具有两类基本形式：单向S-粗集（one direction singular rough sets），单向S-粗集对偶（dual of one direction singular rough sets）； 双向S-粗集(two direction singular rough sets). 单向S-粗集具有单向动态特性，双向S-粗集具有双向动态特性，S-粗集为动态数据挖掘-动态知识发现研究提供了理论支持.史开泉教授提出的S-粗集，开创了粗集研究的新领域，S-粗集的动态特性，在经济系统、管理系统、通信系统、故障检测-识别系统、信息辨识系统、系统状态分析、系统状态识别、系统信息融合、生物医学工程、材料科学与工程、应用数学等领域得到广泛应用。

文献[18]在文献[1]的基础上提出双向变异S-粗集和它的结构.S-粗集与Z. Pawlak粗集的共同特征是以元素的 R-等价类来定义的，S-粗集与Z. Pawlak粗集的主要差异是S-粗集具有动态性，而且S-粗集的动态性又是通过元素迁移体现出来的，由于元素迁移在很多情况下是不确定的，表现出一定的随机性，也就是说不同的元素迁移发生的可能性大小是不同的. 由于S-粗集的随机特征，自然使得双向变异S-粗集也具有概率特征. 在实际问题中，一个系统受到

相关因素的随机干扰是普遍存在的，从双向变异S-粗集的角度来看就是属性迁移具有随机性．因此，作为粗系统分析的理论基础，双向变异S-粗集自然不能忽视属性迁移的随机性，在考虑属性迁移的随机性时，双向变异S-粗集有什么样的结构和特征？这一问题，在有关S-粗集的文献中尚未进行讨论．依据这一背景，本文提出了属性迁移的概率、概率函数的概念，讨论了双向变异S-粗集的概率特征．

第二章　变异 S-粗集

2.1 单向变异 S-粗集

约定：V是有限属性论域，α是V上的有限属性集，$\alpha \subset V$；F是定义在V上的元素迁移，$F=\{f_1, f_2, \cdots, f_m\}$，［$\alpha$］是$\alpha$-属性等价类.

定义 1　称α°是$\alpha \subset V$的单向 S-集合，如果存在$\beta \in V$，$\beta \in \alpha$，$f(\beta)=\alpha' \in \alpha$，而且

$$\alpha^{\circ}=\alpha \cup \{\beta \mid \beta \in V, \beta \overline{\in} \alpha, f(\beta)=\alpha' \in \alpha\} \qquad (1)$$

α^{f}称作$\alpha \subset V$的f-扩张，而且

$$\alpha^{f}=\{\beta \mid \beta \in V, \beta \overline{\in} \alpha, f(\beta)=\alpha' \in \alpha\} \qquad (2)$$

定义 2　设$\alpha^{\circ} \subset V$是属性集$\alpha \subset V$的单向 S-集合，称

$$\begin{aligned}(X^{\circ},F)_{\circ}(\alpha^{\circ}) &= \cup[\alpha] \\ &= \{\alpha \mid \alpha \in V, [\alpha] \subseteq \alpha^{\circ}\}\end{aligned} \qquad (3)$$

是$\alpha^{\circ} \subset V$的下近似；称

$$\begin{aligned}(X^{\circ},F)^{\circ}(\alpha^{\circ}) &= \cup[\alpha] \\ &= \{\alpha \mid \alpha \in V, [\alpha] \cap \alpha^{\circ} \neq \phi\}\end{aligned} \qquad (4)$$

是$\alpha^{\circ} \subset V$的上近似.

这里：$X^{\circ} \subset U$，U 是有限元素论域.

定义 3　集合对 $((X^{\circ},F)_{\circ}(\alpha^{\circ}),(X^{\circ},F)^{\circ}(\alpha^{\circ}))$ 称作单向 S−粗集 $((R,F)_{\circ}(X^{\circ}),(R,F)^{\circ}(X^{\circ}))$ 生成的单向变异 S−粗集，简称单向变异 S−粗集，如果 α−属性等价类族 $\mathrm{Y}[\alpha]$ 是 R−元素等价类族 $\mathrm{Y}[x]_R$ 的单向变异生成.

定义 4　称 $B_{n\alpha}(\alpha^{\circ})$ 是 $\alpha^{\circ} \subset V$ 的 α−边界，而且

$$B_{n\alpha}(\alpha^{\circ})=(X^{\circ},F)^{\circ}(\alpha^{\circ})-(X^{\circ},F)_{\circ}(\alpha^{\circ}) \tag{5}$$

定义 5　称 $As(\alpha^{\circ})$ 是双向变异 S−粗集（$(X^{\circ},F)_{\circ}(\alpha^{\circ})$，$(X^{\circ},F)^{\circ}(\alpha^{\circ})$）生成的副集合，而且

$$As(\alpha^{\circ})=\{\beta \mid \beta \in V, \beta \overline{\in} \alpha, f(\beta)=\alpha' \,\tilde{\in}\, \alpha\} \tag{6}$$

这里：“$\tilde{\in}$”是一个特别的记号，“$\tilde{\in}$”表示 $f(\beta)$ 与 α 的关系具有特征函数 $0<\chi_{\alpha}^{f(\beta)}<1$，它的直接意义是：属性 $\beta \in V$ 在 $f \in F$ 的作用下变成 $f(\beta)=\alpha'$，$f(\beta)$ 不被完全迁入到 α 内．在(2)中，属性 $\beta \in V$ 变成 $f(\beta)=\alpha'$，$f(\beta)$ 与 α 的关系具有特征函数 $\chi_{\alpha}^{f(\beta)}=1$，它的直接意义是：属性 $\beta \in V$ 在 $f \in F$ 的作用下变成 $f(\beta)=\alpha'$，$f(\beta)$ 被完全迁入到 α 内.

属性 $\beta \in V$ 的实际与应用意义：属性集 $\alpha \subset V$ 是人们分析系统时事先确定的，或者说人们已知 α 的存在. 属性 β 是 V 上的新生属性，它的存在人们事先不能确定，属性 β 具有“突发性”，“不可预测性”， 属性 β 广泛地存在于“风险估计”，“预警分析”，“风险管理”，“金融风险分析”等诸多系统中.

由定义 1～5 得到：

命题 1　单向变异 S-粗集是具有单向动态特性的属性集 $\alpha^{\circ} \subset V$ 的粗集.

2.2　双向变异 S−粗集

约定：V 是有限属性论域，α 是 V 上的有限属性集，$\alpha \subset V$；F，$\overline{F}$ 是定义在 V 上的元素迁移，$F = \{f_1, f_2, \cdots, f_m\}$，$\overline{F} = \{\bar{f}_1, \bar{f}_2, \cdots \bar{f}_n\}$，$\mathrm{F} = F \cup \overline{F}$，［$\alpha$］是 α－属性等价类.

定义 6　称 α^* 是 $\alpha \subset V$ 的双向 S−集合，如果存在 $\beta_i \in V$，$\beta_i \,\bar{\in}\, \alpha$，$f(\beta_i) = \alpha_i' \in \alpha$；存在 $\alpha_j \in \alpha, \bar{f}(\alpha_j) = \beta_j \,\bar{\in}\, \alpha$，而且

$$\alpha^* = \hat{\alpha} - \{\alpha \mid \alpha_j \in \alpha, \bar{f}(\alpha_j) = \beta_j \,\bar{\in}\, \alpha\} \qquad (7)$$

$\alpha^{\bar{f}}$ 称作 $\alpha \subset V$ 的 $\bar{f}$−萎缩，而且

$$\alpha^{\bar{f}} = \{\alpha \mid \alpha_j \in \alpha, \bar{f}(\alpha_j) = \beta_j \,\bar{\in}\, \alpha\} \qquad (8)$$

这里：$\hat{\alpha} = \alpha \cup \alpha^{f} = \alpha \cup \{\beta \mid \beta_i \in V, \beta_i \,\bar{\in}\, \alpha, f(\beta_i) = \alpha_i' \in \alpha\}$　　(9)

定义 7　设 $\alpha^* \subset V$ 是属性集 $\alpha \subset V$ 的双向 S−集合，称

$$\begin{aligned}(X^*, \mathrm{F})_{\circ}(\alpha^*) &= \cup[\alpha] \\ &= \{\alpha \mid \alpha \in V, [\alpha] \subseteq \alpha^*\}\end{aligned} \qquad (10)$$

是 $\alpha^* \subset V$ 的下近似，称

$$(X^*, \mathrm{F})^{\circ}(\alpha^*) = \cup[\alpha]$$

$$=\{\alpha \mid \alpha \in V, [\alpha] \cap \alpha^* \neq \phi\} \qquad (11)$$

是 $\alpha^* \subset V$ 的上近似.

这里：$X^* \subset U$，$X^* = \{X \cup \{u \mid u \in U, u \overline{\in} X, f(u) = x \in X\}\} \setminus$

$\{x \mid x \in X, \bar{f}(x) = u \overline{\in} X\}$，$U$是有限元素论域.

定义 8　集合对（$(X^*,\mathrm{F})_{\circ}(\alpha^*)$，$(X^*,\mathrm{F})^{\circ}(\alpha^*)$）称作双向 S-粗集（($R$，F）$_{\circ}(X^*)$,（$R$，F）$^{\circ}(X^*)$)生成的双向变异 S-粗集，简称双向变异 S-粗集，如果α-属性等价类族$\cup[\alpha]$是R-元素等价类族$\cup[x]_R$的双向变异生成.

定义 9　称$B_{n\alpha}(\alpha^*)$是$\alpha^* \subset V$的α-边界，而且

$$B_{n\alpha}(\alpha^*) = (X^*,\mathrm{F})^{\circ}(\alpha^*) - (X^*,\mathrm{F})_{\circ}(\alpha^*) \qquad (12)$$

定义 10　称$As(\alpha^*)$是双向变异 S-粗集（$(X^*,\mathrm{F})_{\circ}(\alpha^*)$, $(X^*,\mathrm{F})^{\circ}(\alpha^*)$）生成的副集合，而且

$$As(\alpha^*) = \{\beta \mid \beta_i \in V, \beta_i \overline{\in} \alpha, f(\beta_i) = \alpha_i' \tilde{\in} \alpha \text{ and } \alpha_j \in \alpha, \bar{f}(\alpha_j) = \beta_j \underset{o/}{\in} \alpha\} \qquad (13)$$

这里："$\underset{o/}{\in}$"是一个特别的记号，"$\underset{o/}{\in}$"表示$\bar{f}(\alpha_j)$与α的关系具有特征函数$-1 < \chi_{\alpha}^{\bar{f}(\alpha_j)} < 0$，它的直接意义是：属性$\alpha_j \in \alpha$在$\bar{f} \in \bar{F}$的作用下变成$\bar{f}(\alpha_j) = \beta_j$，$\bar{f}(\alpha_j)$不被完全迁出到$\alpha$外. 在(8)中，属性$\alpha_j \in \alpha$变成$\bar{f}(\alpha_j) = \beta_j$，$\bar{f}(\alpha_j)$与$\alpha$的关系具有特征函数$\chi_{\alpha}^{\bar{f}(\alpha_j)} = -1$，它的直接意义是：属性$\alpha_j \in \alpha$在$\bar{f} \in \bar{F}$的作用下变成$\bar{f}(\alpha_j) = \beta_j$，$\bar{f}(\alpha_j)$被完全迁出到$\alpha$外.

属性$\alpha_j \in \alpha$的实际与应用意义： 属性α_j是$\alpha \subset V$中可删除的属性，α_j广泛

地存在于“阶段决策”系统中；在一个由多个阶段决策构成的决策系统中，t 阶段决策中的属性$\alpha_j \in \alpha$在 t+1 阶段决策中失效，在 t+1 阶段决策中，属性$\alpha_j \in \alpha$应当从属性集α中删除，$\bar{f}(\alpha_j) = \beta_j \overline{\in} \alpha$，使得 t+1 阶段的决策分析获得简化.

由定义 6～10 得到：

命题 2 双向变异 S-粗集是具有双向动态特性的属性集$\alpha^* \subset V$的粗集.